

*Bioinformatics*, 35(8), 2019, 1380–1387

doi: 10.1093/bioinformatics/bty782

Advance Access Publication Date: 8 September 2018

Original Paper

OXFORD

Systems biology

# Filtering and inference for stochastic oscillators with distributed delays

Silvia Calderazzo<sup>1,2,\*</sup>, Marco Brancaccio<sup>3</sup> and Bärbel Finkenstädt<sup>1,\*</sup>

<sup>1</sup>Department of Statistics, University of Warwick, Coventry CV4 7AL, UK, <sup>2</sup>Division of Biostatistics, German Cancer Research Center, 69120 Heidelberg, Germany and <sup>3</sup>Division of Neurobiology, Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, UK

\*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on April 12, 2018; revised on August 8, 2018; editorial decision on August 30, 2018; accepted on September 6, 2018

## Abstract

**Motivation:** The time evolution of molecular species involved in biochemical reaction networks often arises from complex stochastic processes involving many species and reaction events. Inference for such systems is profoundly challenged by the relative sparseness of experimental data, as measurements are often limited to a small subset of the participating species measured at discrete time points. The need for model reduction can be realistically achieved for oscillatory dynamics resulting from negative translational and transcriptional feedback loops by the introduction of probabilistic time-delays. Although this approach yields a simplified model, inference is challenging and subject to ongoing research. The linear noise approximation (LNA) has recently been proposed to address such systems in stochastic form and will be exploited here.

**Results:** We develop a novel filtering approach for the LNA in stochastic systems with distributed delays, which allows the parameter values and unobserved states of a stochastic negative feedback model to be inferred from univariate time-series data. The performance of the methods is tested for simulated data. Results are obtained for real data when the model is fitted to imaging data on *Cry1*, a key gene involved in the mammalian central circadian clock, observed via a luciferase reporter construct in a mouse suprachiasmatic nucleus.

**Availability and implementation:** Programmes are written in MATLAB and Statistics Toolbox Release 2016b, The MathWorks, Inc., Natick, Massachusetts, USA. Sample code and *Cry1* data are available on GitHub <https://github.com/scalderazzo/FLNADD>.

**Contact:** [s.calderazzo@dkfz-heidelberg.de](mailto:s.calderazzo@dkfz-heidelberg.de) or [b.f.finkenstadt@warwick.ac.uk](mailto:b.f.finkenstadt@warwick.ac.uk)

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics online*.

## 1 Introduction

The time evolution of molecular counts of chemical species in a reaction network is formally described by a Markov jump process (MJP). Interest usually lies in inferring the reaction rates and the unobserved molecule counts of the network's species, given experimental data observed at discrete time-intervals for some or all of the reactants. Biochemical reaction networks are complex and often involve many reactions and chemical species. This is in stark contrast to the fact that only small subsets of species can be observed, albeit indirectly through measurement processes involving e.g. fluorescent

reporter protein imaging. Hence, model reductions of the full reaction network towards less parameter-intensive approaches that can feasibly be estimated from the experimental data are of considerable importance.

The introduction of time-delays, in fixed or distributed form, can approximate the network's species and reaction events which are not of primary interest and thus reduce model complexity (see e.g. [Ananthasubramaniam \*et al.\*, 2014](#); [Heron \*et al.\*, 2007](#); [Korenčič \*et al.\*, 2012](#); [Monk, 2003](#)). Furthermore, it is well known that delays, as well as negative feedback, non-linearity and appropriate

time-scales for the network's reactions are necessary for the onset of oscillations in mathematical models (Cao *et al.*, 2016; Novák and Tyson, 2008). Delays can account for non-observable intermediate species or for time-intervals during which there are no measurements relating to the products of reactions. From the mathematical modelling point of view, this implies that the Markov property holds on a longer time-interval which extends up to an assumed maximum delay time.

A further main challenge in this field is due to the intractability of the transition densities of the MJP. Approximations in continuous state-space are available when suitable assumptions on the system size hold, and in particular, inferential applications have been considered for the chemical Langevin equation (CLE) (Golightly and Wilkinson, 2005, 2010; Heron *et al.*, 2007) and the linear noise approximation (LNA) (Fearnhead *et al.*, 2014; Finkenstädt *et al.*, 2013; Komorowski *et al.*, 2009; Stathopoulos and Girolami, 2013). For a recent detailed review of modelling and inferential methods for stochastic biochemical systems see Schnoerr *et al.* (2017). The CLE aims at matching the infinitesimal mean and variance of the original MJP, while the LNA performs a linearization which leads to tractable Gaussian transition densities. Fearnhead *et al.* (2014) find that in cases where the dynamics are non-linear, parameter inference via the LNA is improved by filtering, i.e. by replacing the mean and variance estimates of the process with their predicted value given the past observations.

The LNA for models with distributed delays has recently been derived by Brett and Galla (2013), but has not been suggested for filtering and inferential purposes, while it appears that a filtering methodology so far has only been suggested for fixed delays (Gopalakrishnan *et al.*, 2011). Here we develop a novel filtering algorithm that is based on the LNA and is generally applicable to stochastic systems comprising distributed delays, with a focus on dynamic state-space models for chemical reaction networks.

We first introduce biochemical reaction networks, their exact mathematical description and the CLE and LNA approximations. We then consider their CLE and LNA approximation in the broader framework of state-space models, and illustrate the LNA updating algorithm in the context of non-delayed systems. We present our novel extension, where the methodology is tested on simulated data and then applied to experimental data. Here we focus on providing an example of a stochastic transcriptional translational feedback loop (TTFL) to describe the expression dynamics of the circadian gene *Cry1*. The methodology is used to infer parameters of the TTFL of *Cry1* from experimental time-series data observed in a mouse suprachiasmatic nucleus (SCN) tissue (Brancaccio *et al.*, 2013).

## 2 Materials and methods

### 2.1 Reaction networks and their approximations

Consider a reaction network defined by a set of chemical species participating in a set of chemical reactions. Let  $p$  and  $r$  denote the total number of species and reactions, respectively. The state of the process  $X(t)$  can be defined as (Anderson and Kurtz, 2011)

$$X(t) = X(0) + SU \left( \int_0^t b(X(s), c) ds \right),$$

where the vector of random variables  $X(t) = (X_1(t), \dots, X_p(t))^T$  defines the number of molecules at time  $t$  of the species participating in the reaction network,  $U(\cdot)$  is a vector of  $r$  independent

inhomogeneous Poisson processes counting the occurrence of the  $r$  reactions and its argument defines its mean,  $S$  is the  $p \times r$  stoichiometry matrix whose elements  $s_{i,k}$  denote the difference in the number of molecules of the  $i$ -th species produced and consumed by the  $k$ -th reaction. The vector  $b(X(s), c) = (b_1(X(s), c_1), \dots, b_r(X(s), c_r))^T$  contains the reaction hazards where  $c_k$  denotes the rate constant of the  $k$ -th reaction. The resulting process  $X(t)$  is a MJP in continuous time and discrete state-space, and its Kolmogorov's forward equation, or chemical master equation (CME) (Gillespie, 1992), provides an exact description of the system. As it can only be explicitly solved in rare cases (see the review in McQuarrie, 1967), approximations help to overcome the need for highly computationally demanding inferential procedures. The CLE, or diffusion approximation, exploits the multivariate normal approximation of the vector of independent Poisson random variables  $U$ . The approximation holds under suitable assumptions concerning the number of reaction occurrences, and leads to the stochastic differential equation form (Anderson and Kurtz, 2011)

$$dX(t) = g(X(t), c)dt + \sqrt{A(X(t), c)}dB(t), \quad (1)$$

where  $dB(t)$  is a  $p$ -dimensional Wiener process, and

$$g(X(t), c) = Sb(X(t), c)$$

$$A(X(t), c) = S \text{diag}\{b(X(t), c)\}S^T.$$

The drawback of the CLE approximation is that explicit solutions for the transition densities are again rare (see e.g. Wilkinson, 2012).

Gaussian transition densities can be obtained with the LNA (Kurtz, 1972; Van Kampen, 1992) which also exploits the approximate normality of the underlying Poisson process, but replaces the hazard function by its first order Taylor expansion about the deterministic limit of the process, thus effectively eliminating non-linearities. Assuming  $X(0) \sim \mathcal{N}(\rho(0), P(0))$ , where  $\rho$  denotes the deterministic limit, it can be shown that

$$X(t) \sim \mathcal{N}(\rho(t), P(t))$$

such that the mean and the variance are the solutions of

$$d\rho(t) = g(\rho(t), c)dt \quad (2)$$

$$dP(t) = J_g(\rho(t), c)P(t)dt + P(t)^T J_g(\rho(t), c)^T dt + A(\rho(t), c)dt, \quad (3)$$

where  $J_g$  is the Jacobian of  $g(\cdot)$  (see e.g. Anderson and Kurtz, 2011, for a rigorous derivation). The LNA matches exactly the first two moments provided by the CME for systems including reactions up to the first order, as well as for a subset of systems including second-order reactions (Grima, 2015). In the other cases, the LNA also approximates the mean and variance, as they depend on the higher moments (see e.g. Gillespie and Golightly, 2016; Grima, 2012, 2015).

### 2.2 Filtering and inference for the LNA

Inference from experimental data requires the formulation of a measurement equation. If light intensities generated by fluorescent reporter constructs are recorded, we may assume that observations are proportional to the unobserved molecular abundance and subject to additive Gaussian measurement error with unknown variance (see Finkenstädt *et al.*, 2013; Folia and Rattray, 2017; Hey *et al.*, 2015; Komorowski *et al.*, 2009)

$$Y_t = FX_t + \epsilon_t, \tag{4}$$

where  $\epsilon_t \sim MVN(0, \Sigma_\epsilon)$ ,  $F$  is a  $q \times p$  matrix (where  $q$  is the dimension of  $Y$  and  $p$  the dimension of  $X$ ). The Markov structure of the process together with the measurement process in (4) formally leads to a state-space model, which provides a framework for inference. Let  $\psi$  denote the set of all parameters, namely the vector of reaction rates  $c$ , the variance matrix  $\Sigma_\epsilon$  and the scale parameters in  $F$ . For random variables  $X$  and  $Y$ , let  $\pi(x)$  define the probability density function of  $X$ , and  $\pi(x|y)$  the conditional probability density function of  $X$  given  $Y$ . Furthermore, let  $Y_{0:T} = Y_0, Y_{\Delta_t}, \dots, Y_T$  and  $X_{0:T} = X_0, X_{\delta_t}, \dots, X_T$ , where  $\Delta_t$  is the time-interval of the observations,  $\delta_t$  some suitable discretization interval for the unobserved states dynamics and  $T$  is the total observation time. Inference typically involves estimating the model parameters  $\psi$  and the hidden states  $X_{0:T}$ . According to Bayes' theorem, their joint posterior distribution is

$$\pi(\psi, x_{0:T}|y_{0:T}) = \frac{\pi(y_{0:T}|x_{0:T}, \psi)\pi(x_{0:T}|\psi)\pi(\psi)}{\pi(y_{0:T})}.$$

If the focus of inference lies on estimating the model parameters, their marginal posterior distribution  $\pi(\psi|y_{0:T})$  can be obtained by integrating out the hidden states, while inference on the hidden states can be performed by computing the smoothing density  $\pi(x_{0:T}|y_{0:T}, \psi)$  (see e.g. Doucet and Johansen, 2009; Wilkinson, 2012 and the Supplementary Section 1). In Gaussian dynamic linear models, all densities involved are Gaussian and the marginal likelihood  $\pi(y_{0:T}|\psi)$  is available in a closed form provided by the well-known Kalman filtering methodology (Kalman, 1960). However, in reaction networks leading to the CLE representation of (1), Gaussianity is lost due to the dependence of the diffusion term  $A(\cdot)$  on the system state. Furthermore, when the hazard function  $b(\cdot)$  is non-linear, i.e. for reactions of order greater than one, only an approximate estimate of the mean and variance of the process is generally available.

In the filtering literature, non-linear and/or non-Gaussian systems have been approached with a variety of methods, e.g. the extended and second-order extended Kalman filter (see e.g. Jazwinski, 2007; Särkkä, 2013; Singer, 2002), the unscented Kalman filter (Julier and Uhlmann, 1997) and particle filters (see e.g. Andrieu et al., 2010; Doucet and Johansen, 2009; Golightly and Wilkinson, 2011). Here we focus on the extended Kalman filter for time-continuous unobserved states, namely the extended Kalman-Bucy filter (EKBF) (see e.g. Kulikov and Kulikova, 2014; Singer, 2002 and references therein). Such choice can be motivated by its link to the LNA as follows.

Assume, for ease of notation, that the parameters  $\psi$  are set to some fixed value. In practice, parameter estimation can be performed via e.g. a Bayesian Markov chain Monte Carlo (MCMC) algorithm. The EKBF performs an update (restart) of the LNA mean and variance estimates at each observation time-point. Suppose  $\rho(t) = E[X(t)|y_{0:t}]$  and  $P(t) = \text{Var}[X(t)|y_{0:t}]$  in (2) and (3), respectively. Assuming that  $\pi(x_i|y_{0:t})$  is approximately Gaussian, linearity of the LNA and of the measurement process implies that  $\pi(x_{t+\Delta_t}|y_{0:t})$  and  $\pi(y_{t+\Delta_t}|y_{0:t})$  are Gaussian, the latter with mean  $F\rho_{t+\Delta_t}$ , and variance  $FP_{t+\Delta_t}F^T + \Sigma_\epsilon$ . Estimates of the mean and variance of  $\pi(x_{t+\Delta_t}|y_{0:t+\Delta_t})$  are obtained by a Kalman filtering step

$$\rho_{t+\Delta_t}^* = \rho_{t+\Delta_t} + C(y_{t+\Delta_t} - F\rho_{t+\Delta_t}) \tag{5}$$

$$P_{t+\Delta_t}^* = P_{t+\Delta_t} - CFP_{t+\Delta_t}, \tag{6}$$

where  $C = P_{t+\Delta_t}F^T(FP_{t+\Delta_t}F^T + \Sigma_\epsilon)^{-1}$  is the adaptive coefficient. The current mean and variance estimates  $\rho_{t+\Delta_t}$  and  $P_{t+\Delta_t}$  are then replaced

by their optimal estimates  $\rho_{t+\Delta_t}^*$  and  $P_{t+\Delta_t}^*$ , and computations are iterated up to  $T$ . The update Equations (5) and (6) can be derived from approximate Gaussianity of the joint distribution  $\pi(x_{t+\Delta_t}, y_{t+\Delta_t}|y_{0:t})$ , and the subsequent conditioning upon  $Y_{t+\Delta_t}$ . For further details see e.g. Jazwinski (2007), Särkkä (2013) and the Supplementary Section 2.

### 2.3 Extension to systems with distributed delays

Distributed delays can be introduced in the hidden state equation to account for a dependency on a (finite or infinite) collection of past states that are arbitrarily distant in time.

We focus on systems where a set of intermediate transformations of the species of interest can be well approximated by the Goodwin oscillator ordinary differential equations (ODEs) (Goodwin, 1965), which can be explicitly solved. The resulting system is a reduced reaction network, in which the hazard accounting for the transcriptional process receives as an input its past expression level, weighted according to the delay distribution. As a Hill function is assumed for transcription, this leads to a delay entering non-linearly in the transcriptional hazard. A formal derivation of the model is provided in the Supplementary Section 3. Further motivation or pursuing this modelling framework is provided in Section 3.1.

Brett and Galla (2013) derive a CLE and a (non-restarted) LNA for a modelling framework in which the delays are included linearly in the reaction hazards. Such scenario may arise under the assumption that the product of a reaction is not available for a non-negligible random time-interval, as well as under alternative modelling assumptions e.g. for an alternative target state in the Goodwin oscillator system.

To derive the LNA approach for our modelling scenario, we follow an approach analogous to Brett and Galla (2013), and divide the reactions into two groups: the set of  $w$  reactions with distributed delays, with stoichiometry  $S_d$  of dimension  $p \times w$  and hazard vector  $b_d$  of length  $w$ , and the set of  $z$  reactions not involving delayed species, with stoichiometry matrix  $S_{\bar{d}}$  of dimension  $p \times z$ , and hazard vector  $b_{\bar{d}}$  of length  $z$ . We have  $S = [S_d|S_{\bar{d}}]$  and for the hazards

$$b_d \left( \int_{-\infty}^t X(s) \cdot K(t-s) ds \right) = \begin{bmatrix} b_{d1} \left( \int_{-\infty}^t X(s) \cdot K(t-s) ds \right) \\ \vdots \\ b_{dw} \left( \int_{-\infty}^t X(s) \cdot K(t-s) ds \right) \end{bmatrix},$$

where  $K(\cdot) = [K_1(\cdot), \dots, K_p(\cdot)]^T$  is a vector of delay densities, one for each  $X_i$ . If  $X_i$  is not delayed,  $K_i$  has a point mass density at 0. The hazard vector of the non-delayed reactions is simply  $b_{\bar{d}}(X(t)) = [b_{w+1}(X(t)), \dots, b_{w+z}(X(t))]^T$ . We have dropped the dependence on the reaction rates  $c$  for ease of notation. It is practical to assume that the delay distribution is truncated at some maximum delay time  $\tau_m$ . Given the knowledge of the past states of the system, the hazards are deterministically defined and the CLE for the reaction network is

$$dX(t) = g(X(t))dt + f \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right) dt + \sqrt{l(X(t)) + q \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right)} dB(t), \tag{7}$$

where  $dB(t)$  is as usual a  $p$ -dimensional Wiener process, and

$$g(\cdot) = S_{\bar{d}}b_{\bar{d}}(X(t))$$

$$f(\cdot) = S_d b_d \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right)$$

$$l(\cdot) = S_d \text{diag}[b_d(X(t))] S_d^T$$

$$q(\cdot) = S_d \text{diag} \left[ h_d \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right) \right] S_d^T.$$

The full state-space model is hence given by the continuous approximation for the dynamics of the unobserved molecule counts in (7) together with the measurement Equation (4).

The derivation of the filtering algorithm poses two difficulties: (i) the linearization of the functions incorporating the delayed species, and (ii) the need to update, at each observation time  $t$ , all estimates in the past until time  $t - \tau_m$ . Problem (i) is addressed by a first order Taylor expansion of  $g(X(t))$  and  $l(X(t))$  about  $\rho(t)$ , and of  $f\left(\int_{t-\tau_m}^t X(s) \cdot K(t-s) ds\right)$  and  $q\left(\int_{t-\tau_m}^t X(s) \cdot K(t-s) ds\right)$  about  $\int_{t-\tau_m}^t \rho(s) \cdot K(t-s) ds$ . The expansions are plugged into the mean and variance equations, thus allowing their propagation under linearity. A detailed derivation of the mean and variance equations is provided in the Supplementary Section 4. Optimal estimates of  $\rho(t)$  and  $P(t)$  for  $t \in [0, \tau_m]$ , required for the initialization of the algorithm, are generally not available in practice and require further modelling.

Problem (ii) can be approached by performing a Kalman update analogous to (5) and (6). Assuming approximate Gaussianity of the joint distribution  $\pi(x_{t+\Delta_t-\tau_m:t+\Delta_t}, y_{t+\Delta_t} | y_{0:t})$  and conditioning on  $Y_{t+\Delta_t}$ , implies that  $\pi(x_{t+\Delta_t-\tau_m:t+\Delta_t} | y_{0:t+\Delta_t})$  is approximately Gaussian with mean  $\rho_{t+\Delta_t-\tau_m:t+\Delta_t}^*$  and variance  $P_{t+\Delta_t-\tau_m:t+\Delta_t}^*$ , where

$$\rho_{t+\Delta_t-\tau_m:t+\Delta_t}^* = \rho_{t+\Delta_t-\tau_m:t+\Delta_t} + C(y_{t+\Delta_t} - F\rho_{t+\Delta_t})$$

$$P_{t+\Delta_t-\tau_m:t+\Delta_t}^* = P_{t+\Delta_t-\tau_m:t+\Delta_t} - CFP_{t+\Delta_t,t+\Delta_t-\tau_m:t+\Delta_t}$$

with coefficient of adaptation  $C$

$$C = P_{t+\Delta_t-\tau_m:t+\Delta_t} F^T (FP_{t+\Delta_t} F^T + \Sigma_\epsilon)^{-1}.$$

At  $t + \Delta_t$  we therefore update, conditional on  $y_{0:t+\Delta_t}$ , the unobserved state mean and variance backwards until  $t + \Delta_t - \tau_m$ . At the end of the observation time  $T$  such updates effectively provide a ‘partial’ smoothing density, as the moments of  $x_{T-\tau_m:T}$  are conditional on  $y_{0:T}$ , those of  $x_{T-\tau_m-\Delta_t:T-\tau_m}$  are conditional on  $y_{0:T-\Delta_t}$ , etc.

The complexity introduced by the distributed delay comes at higher computational cost. When the dimension of  $X$  and, more crucially, the number of unobserved states included in the maximum delay time is large, the algorithm can be significantly slower than in the non-delayed case. This point is further investigated in Section 3.2.

### 3 Application

#### 3.1 Model

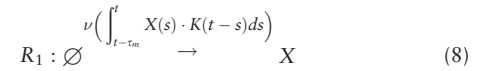
We consider a stochastic transcriptional and translational feedback loop (TTFL) for a circadian clock that is represented by the self-inhibition of transcription after a Gamma distributed delay time  $\tau_p$ . The delay accounts for nuclear export, protein synthesis and nuclear import (see the Supplementary Section 3 for a formal derivation). We assume maximum delay time,  $\tau_m = 30$  h, such that possible dependence on past states is limited to just over a day, i.e. the cycle length of the system. Furthermore, we assume that a Hill function can approximate the relationship between the amount of available inhibitor and the transcriptional rate output, i.e.

$$\nu \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right) = \frac{R_{max}}{1 + \left( \frac{\int_{t-\tau_m}^t X(s) \cdot K(t-s) ds}{K_{pc}} \right)^n},$$

where  $\nu$  is thus the transcription function. The parameter  $R_{max}$  is the maximum achievable transcription rate,  $n$  is the Hill coefficient and

is related to the number of binding sites present in the promoter region of the regulated gene and  $K_{pc}$  is called the dissociation coefficient, or threshold, and represents the amount of input required to decrease the output of the transcriptional function by 50%. The Hill function can be formally derived under the assumption that binding and unbinding reactions of the inhibitor to the promoter happen at a fast time-scale, if compared to the time-scale of the transcriptional reaction (see e.g. Tkačik and Walczak, 2011).

The elimination of the intermediate species by means of the distributed delay and the use of a Hill type transcription function, leads to a reduced reaction network which consists of the following two reactions



where  $X$  denotes the mRNA,  $K(\cdot)$  is the truncated Gamma probability density function and  $\mu$  is the mRNA degradation rate. Note that a key assumption has been made by considering the transcriptional process as the relevant source of intrinsic stochasticity.

The continuous state-space approximation of this model, combined with a measurement equation, is

$$Y_t = \kappa \int_{t-\Delta_t}^t X(s) ds + \epsilon_t$$

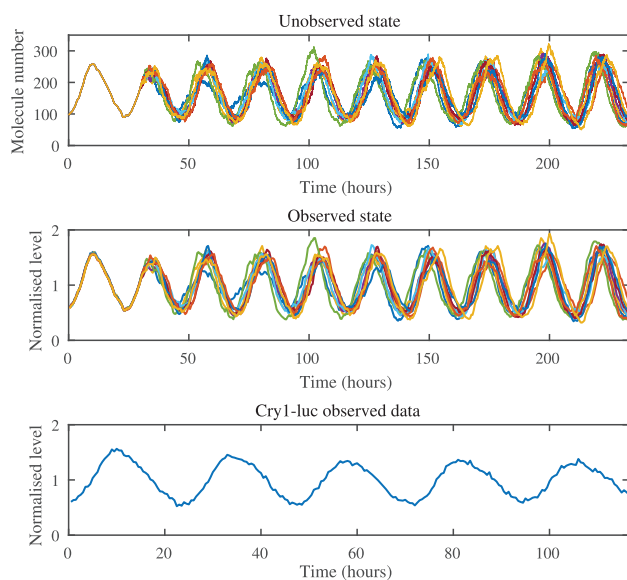
$$dX(t) = \left[ \nu \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right) - \mu X(t) \right] dt + \sqrt{\nu \left( \int_{t-\tau_m}^t X(s) \cdot K(t-s) ds \right) + \mu X(t)} dB(t),$$

where  $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ , and  $\kappa$  is the proportionality factor relating the unobserved molecular process to the light intensity. Here, the measurement Equation (4) has been modified to reflect a measurement process which involves integration of the light signal (see also Folia and Rattray, 2017). The model requires the specification of the initial condition, i.e. the mean and variance of  $\pi(x_{0:\tau_m} | y_{0:\tau_m})$ . To keep the dimensionality of the parameter space to a minimum, we assume a step function for the transcription rate  $\nu$ , thus eliminating the dependence on past mRNA (Hey *et al.*, 2015; Jenkins *et al.*, 2013). Further details are provided in the Supplementary Section 5.

#### 3.2 Simulation study

We use the stochastic simulation algorithm (SSA) (Gillespie, 1977) to simulate approximately the dynamics of this reaction network where the hazard for the reaction in (8) is computed by evaluating the integral accounting for the delay up to the maximum delay time, each selected reaction is then assumed to take place immediately. As initial condition, we adopt the first 30 h of real observations for the *Cry1-luc* imaging data, which was rescaled, aggregated and detrended. Values of mRNA are simulated and stored at fixed time-intervals of duration of length 0.01 h, and, to mimic the integration of camera light, are summed over 0.5 h, divided by their mean level, and corrupted with normal measurement error for two chosen levels of signal to noise ratio, i.e. 20 and 100. Figure 1 shows the simulated time-series for 10 replicates of the simulation algorithm for signal to





**Fig. 1.** Top: SSA-type simulations for the reactions in (8) and (9). Center: molecule counts are rescaled by their mean level, integrated over 0.5 h and corrupted with measurement error. The assumed levels of signal to noise ratio is 100. Bottom: experimental *Cry1-luc* imaging time-series, aggregated, detrended and normalized

noise ratio 100. [Supplementary Figure S1](#) shows the simulated trajectories for signal to noise ratio 20.

### 3.2.1 Filtering performance

We start by verifying the performance of the filtering methodology by setting the parameters to their true values. Computational practice requires some time-discretization of the ODEs involved. We adopt the Euler discretization method assuming  $\delta_t$  to control the step-size of the approximation. [Supplementary Figure S2](#) provides a graphical illustration of the filter performance for  $\delta_t = 0.1$  on one sample simulated time-series. We note overall a good precision and, as expected, a decreased variability in the smoothing distribution. To compare the behaviour of the filter for different values of  $\delta_t$ , we compute the empirical coverage of both the predictive and the partial smoothing densities at level 95%. Results for the two levels of measurement error are provided in [Supplementary Table S1](#). We observe an improved precision, i.e. a decreased discrepancy between nominal and empirical coverage, as  $\delta_t$  approaches its true value of 0.01 h, and particularly for the partial smoothing density under the lower measurement error scenario. The improvement is slower for values of  $\delta_t$  below 0.1 h, at which coverages for all cases are already between 94 and 96%. Furthermore, any improved coverage comes at higher computational cost. [Supplementary Table S2](#) provides average running times of the filter. We observe that the rate of increase of the computational expense increases as  $\delta_t$  decreases, where the computation of the covariance of the unobserved states accounts for a significant proportion of the total running time. We note that alternative ODE solvers may be employed, and that the balance between precision and computational costs needs to be assessed in each case.

### 3.2.2 Inference performance

We study the performance of the proposed filter for the purpose of Bayesian inference by designing a MCMC algorithm and applying it to the last five cycles of the simulated data ([Fig. 1](#)) with the aim of

retrieving the values of the parameters used for the simulations. The MCMC scheme details are given in the [Supplementary Section 6](#). Based on our results above and to balance the trade-off between realistic running times and filtering performance we choose to work with values  $\delta_t \geq 0.1$  h. Visual investigation of the univariate log-likelihoods (see [Supplementary Figs S3 and S4](#)) suggests a minor effect of the choice of  $\delta_t$  on most of the parameters involved, except  $\sigma_\epsilon$  and, more marginally,  $\kappa$ . This result motivates the design of a delayed acceptance MCMC algorithm ([Christen and Fox, 2005; Golightly et al., 2015](#)), which allows to exploit the fast likelihood computation provided by the filter for  $\delta_t = 0.5$  h to explore the parameter space, but finally accepts values according to the likelihood provided by the filter for  $\delta_t = 0.1$  h.

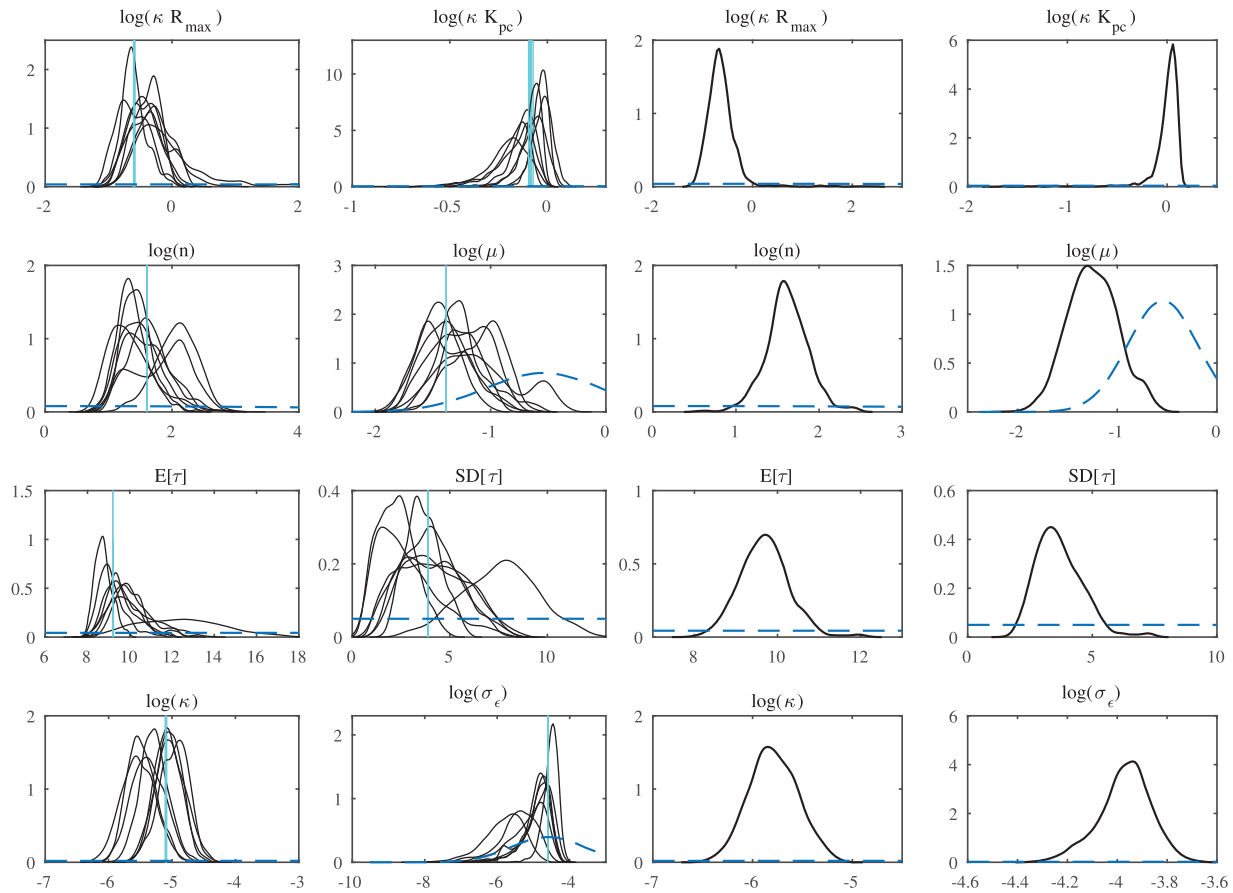
For more efficient sampling we reparameterize the model by moving  $\kappa$  from the observation to the state equation. We generally adopt a  $\mathcal{N}(0, 10^2)$  prior density for all model parameters on the logarithmic scale, with the exceptions of the Hill coefficient, where we assume  $\log(n)$  to have prior  $\mathcal{N}(\log(1.5), 5^2)$  and the degradation rates during and after the initial condition,  $\log(\mu_0)$  and  $\log(\mu)$ , for which we specify a  $\mathcal{N}(\log(0.58), 0.5^2)$  prior. The initial condition dispersion coefficient  $\log(\beta)$ , variance  $\log(\beta \text{Var}[X_0])$  and  $\log(\kappa)$  are assumed to have a  $\mathcal{N}(0, 20^2)$  prior (see [Supplementary Section 5](#) for further details on the initial condition model specification). An informative prior centred at the true simulation value and with unit SD is assumed for the measurement error SD,  $\log(\sigma_\epsilon)$ . Finally, the delay mean and SD are assigned  $U(0, 23)$  and  $U(0, 20)$  prior densities, respectively, both expressed in hours. This can be justified by assuming that the cellular product of the previous circadian cycle is feeding into the dynamics of the next cycle. Hence, most priors have been set to be diffuse within biologically realistic ranges. A certain amount of prior information seems to be required to robustify the inferential process. In particular, prior information on the reporter protein half-life is provided by [Yamaguchi et al. \(2003\)](#) although here we assume a larger SD to cope with the approximate nature of our model in particular as the reporter process is not explicitly modelled. An informative prior for the SD of the measurement error has been deduced by measurements from an additional light channel (M.Unosson, personal communication) where we adopt again a larger variance to indirectly account for processes not explicitly modelled. The initial conditions for the parameter chains are randomly drawn from the prior densities.

[Figure 2](#) shows the estimated posterior densities of the model parameters for the simulated data seen in [Figure 1](#). Two chains have been excluded due to lack of convergence, but inference seems overall satisfactory for the remaining chains. The results for the signal to noise ratio 20 scenario are provided in [Supplementary Figure S5](#), and lead to analogous conclusions.

We have also investigated filter coverage and inference performance for increasing degrees of non-linearity by varying the Hill coefficient and setting it = 7, 9, 11, and found that the estimation results are overall consistent with the results shown for  $n = 5$  (see [Supplementary Tables S3–S5](#) and [Figs S6–S11](#)). We postulate that this is due to the use of the filtering step in the LNA, which seems to effectively counteract the effect of the increasing error in the LNA prediction that can be expected for higher degrees of non-linearity.

### 3.3 Inference for the circadian feedback loop in *Cry1*

The available data are time-series for the circadian gene *Cry1*, observed by mean of a transcriptional luciferase reporter construct *Cry1-luc* in a mouse SCN tissue over five days ([Brancaccio et al., 2013](#)) (see bottom panel of [Fig. 1](#)). Light intensities are recorded by



**Fig. 2.** Left two columns: results for simulation study. Kernel densities' estimates of the model parameters posterior densities, excluding the parameters of the initial condition.  $E[\tau]$  and  $SD[\tau]$  denote the mean and SD of the delay density  $K$ . The prior density is shown as a dashed line while the vertical line marks the true value. Results for the last five cycles of the simulated data shown in the central panel of Figure 1 (two chains are excluded due to non-convergence). Right two columns: results for observed *Cry1-luc* shown in bottom panel of Figure 1. Same notations and definitions as in panels on left

EM-CCD Camera (Hamamatsu) with an exposure time of 0.5 h, i.e. the image is the result of the photons hitting the camera in 0.5 h intervals (see Brancaccio *et al.*, 2013). A mean trend due to consumption of the luciferin substrate is accounted for by dividing the observations by a time-varying proportionality factor which assumes a linear decrease of  $\sim 30\%$  over five days. We apply the proposed model to a sample location in the SCN where the data are aggregated over a  $2 \times 2$  pixel box which is broadly comparable to the size of a cell. Figure 2 provides the estimated posterior densities of the model parameters for the experimental data. Their interpretation has to be considered in the light of the fact that we have represented a complex genetic network of the circadian clock, which consist of several interwoven TTFs involving about 15 clock genes (Dibner and Schibler, 2015) by a delayed feedback loop fitted to imaging data of a single gene involved. The main achievement here over previous work is that we have provided the inferential methodology to identify such a model from time-series imaging data which can potentially serve as a realistic surrogate model to be fitted in a longitudinal or spatial fashion to quantify and compare intrinsic noise and to study between-cell variability. Of particular interest are the parameters associated with the delay distribution, where the mean influences the period of the oscillations, and the variance tunes the temporal precision of the clock period. The mean delay time here is estimated to be around 9.67 h [95% HPDI (8.46, 10.80)] which, although not statistically significant, is not far away from Korenčić *et al.* (2012) who find that a value of 8.25 h achieves

circadian periodicity in a deterministic feedback loop model with discrete delay for *Per2* self-inhibition. Furthermore, we estimate a value of around 3.56 h [95% HPDI (2.01, 5.39)] for the SD of the delay distribution. We note that the inferred posterior distribution of the parameter is concentrated away from 0, which would correspond to a fixed delay, suggesting that the assumption of a distributed delay is indeed tenable. We also hypothesize that this estimate may be relatively small as the experimental data are from the SCN, which is known to be the main pacemaker of the mammalian clock, but may be spatially varying over the SCN. Such further questions can be addressed on the basis of the TTFL modelling and inference approach proposed here.

The model fit is checked through inspection of the data posterior predictive distribution and the standardized residuals, where we obtain samples using a thinned set of parameter samples from the MCMC algorithm. Results are displayed in Supplementary Figure S12. The normal  $q$ - $q$  plot reveals a slightly heavy upper tail, although this result does not seem of major concern nor to be due to a systematic under-performance on the filter, as also investigated in further yet unpublished work. Visual investigation and model diagnostics reveal no residual 24-h periodicity in the residuals after the first 30 h of initial condition (which we recall is only instrumental into obtaining a first estimate of the mean and covariance matrix), which indicates that the circadian oscillation is well explained by the model. However, a non-negligible residual periodicity of around 12 h can still be found. As pointed out to us by an anonymous

reviewer, one possible explanation is that the model may not be capturing part of the non-linearity observed in the data. This may be caused by processes which are not explicitly included in the model such as the influence of inter-cellular signalling. The latter is relevant for *Per* transcription, and therefore for PER protein, which forms the PER/CRY complex effectively repressing *Cry1* transcription. The transcription function assumed above can only account for repression by means of a single transcription factor. It may also be of relevance that [Zhu et al. \(2017\)](#) recently found a cell-autonomous 12 h clock in mammals.

## 4 Discussion

In this work we have addressed filtering and inference for state-space models with distributed delays, with a special focus on models arising from stochastic biochemical networks with stochastic oscillatory behaviour. The methodology is derived with focus on a stochastic self-inhibitory feedback loop with distributed delay, noting that this kind of model is known to achieve a substantial model reduction of complex gene networks and is of particular interest for modelling oscillatory molecular clocks. As such complex networks are never fully observed, model reduction is also important to facilitate inference from experimental data. The introduction of a distributed delay, as compared to a fixed delay, is beneficial as it provides a more realistic description of the process. The type of oscillatory processes here considered can indeed possess a certain degree of variability in their period that a fixed delay would not be able to characterize. We have extended the LNA and EKBF to provide a methodology which allows for sequential computation of the likelihood in such models. The resulting likelihood has a closed form and can be incorporated in a Bayesian MCMC algorithm for parameter inference. The performance of the methodology is tested on simulated data and first real results are shown here for an experimental *Cry1-luc* time-series from a mouse SCN ([Brancaccio et al., 2013](#)).

The closest approach to date is perhaps the work of [Heron et al. \(2007\)](#), which performs inference on a closely related dynamical model for gene expression, where inference is based on the CLE description of the process. In contrast to our methodology, their approach does not allow to obtain a closed-form likelihood and requires sampling the parameters conditional on a sampled path of the unobserved states in the MCMC algorithm and vice-versa, rendering the inference problem very high-dimensional. As the parameters and unobserved states trajectories tend to be strongly correlated, sampling is likely to be less efficient ([Golightly and Wilkinson, 2011](#)).

It should be noted that moment-closure approximations (MA) represent an inferential approach of comparable computational cost which can also be of interest for our scenario. MA provide ODEs for the time evolution of the approximate moments of the process by truncation of higher order moments, and have been applied for inferential purposes in e.g. [Zechner et al. \(2012\)](#) and [Fröhlich et al. \(2016\)](#). The method can lead to physically implausible results, such as negative mean and variances, but [Schnoerr et al. \(2017\)](#) have recently shown an improved precision over the LNA for some simulation scenarios.

Alternative specifications for the transcription function can be investigated, e.g. [Kim et al. \(2014\)](#) assume that the inhibitory processes arise as a consequence of sequestration of the activator by the inhibitor. To date, modelling of the TTFL of circadian genes in the mammalian clock is either based on deterministic approaches, possibly comprising a simplification using a delay, or on stochastic

descriptions which have not yet attempted statistical parameter inference from experimental time-series data (see [Ananthasubramaniam et al., 2014](#); [DeWoskin et al., 2015](#); [Gonze et al., 2005](#); [Kim and Forger, 2012](#); [Korenčič et al., 2012](#); [Relógio et al., 2011](#)). Our approach extends over existing mathematical modelling in that it provides a novel stochastic inferential approach for such system.

We note that our approach does not specifically take into account the fact that we observe reporter protein rather than the protein of interest, which is possibly a strong but necessary simplification. Inference is challenging due to the strong parameter correlation structure, where we hope to show in future work that joint estimation of the parameters for data at many more spatial locations across the SCN may substantially aid the inferential process.

The ability of the proposed methodology to address intrinsic stochasticity is of particular importance in models of biochemical oscillators. Indeed, the idea that noisy systems are more easily entrained to an external input has been investigated both theoretically ([Steuer et al., 2003](#)) and experimentally ([Ko et al., 2010](#)). The latter have studied the role of intrinsic noise in the SCN, and provided experimental evidence from a *Bmal1*-null mutant mouse that noise and extracellular signalling are sufficient to produce oscillations when the TTFL is disrupted. Further insight into the stochastic biochemical oscillators and their synchronization may be achieved by performing inference on the spatial distribution of the model parameters over the SCN. The methodology presented here will form the basis to be able to perform such an analysis.

## Acknowledgements

We thank M. Hastings, D. Rand, G. Minas and M. Unosson for stimulating discussions on the methodology and further insights into the biology. We are grateful to the reviewers for their helpful comments which have significantly improved the manuscript.

## Funding

This work was supported by the Engineering and Physical Sciences Research Council UK (EP/D002060/1); and the Medical Research Council UK (MR/M013170/1 and core funding MC\_U105170643).

*Conflict of Interest:* none declared.

## References

- Ananthasubramaniam, B. et al. (2014) Timing of neuropeptide coupling determines synchrony and entrainment in the mammalian circadian clock. *PLoS Comput. Biol.*, **10**, e1003565.
- Anderson, D.F. and Kurtz, T.G. (2011) Continuous time Markov Chain models for chemical reaction networks. In: *Design and Analysis of Biomolecular Circuits*. Springer, New York, pp. 3–42.
- Andrieu, C. et al. (2010) Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Series B Stat. Methodol.*, **72**, 269–342.
- Brancaccio, M. et al. (2013) A Gq-Ca<sup>2+</sup> axis controls circuit-level encoding of circadian time in the suprachiasmatic nucleus. *Neuron*, **78**, 714–728.
- Brett, T. and Galla, T. (2013) Stochastic processes with distributed delays: chemical Langevin equation and linear-noise approximation. *Phys. Rev. Lett.*, **110**, 250601.
- Cao, Y. et al. (2016) Elements of biological oscillations in time and space. *Nat. Struct. Mol. Biol.*, **23**, 1030–1034.
- Christen, J.A. and Fox, C. (2005) Markov chain Monte Carlo using an approximation. *J. Comput. Graph. Stat.*, **14**, 795–810.

- DeWoskin, D. *et al.* (2015) Distinct roles for GABA across multiple timescales in mammalian circadian timekeeping. *Proc. Natl. Acad. Sci. USA*, **112**, E3911–E3919.
- Dibner, C. and Schibler, U. (2015) Circadian timing of metabolism in animal models and humans. *J. Intern. Med.*, **277**, 513–527.
- Doucet, A. and Johansen, A.M. (2009) A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of Nonlinear Filtering*, **12**, 656–704.
- Fearnhead, P. *et al.* (2014) Inference for reaction networks using the linear noise approximation. *Biometrics*, **70**, 457–466.
- Finkenstädt, B. *et al.* (2013) Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: an application to single cell data. *Ann. Appl. Stat.*, **7**, 1960–1982.
- Folia, M.M. and Rattray, M. (2017) Trajectory inference and parameter estimation in stochastic models with temporally aggregated data. *Stat. Comput.*, **28**, 1–20.
- Fröhlich, F. *et al.* (2016) Inference for stochastic chemical kinetics using moment equations and system size expansion. *PLoS Comput. Biol.*, **12**, e1005030.
- Gillespie, C.S. and Golightly, A. (2016) Diagnostics for assessing the linear noise and moment closure approximations. *Stat. Appl. Genet. Mol. Biol.*, **15**, 363–379.
- Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
- Gillespie, D.T. (1992) A rigorous derivation of the chemical master equation. *Physica A Stat. Mech. Appl.*, **188**, 404–425.
- Golightly, A. *et al.* (2015) Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Stat. Comput.*, **25**, 1039–1055.
- Golightly, A. and Wilkinson, D.J. (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, **61**, 781–788.
- Golightly, A. and Wilkinson, D.J. (2010) Markov chain Monte Carlo algorithms for SDE parameter estimation. In: *Learning and Inference for Computational Systems Biology*. MIT Press, Cambridge, MA, USA, pp. 253–276.
- Golightly, A. and Wilkinson, D.J. (2011) Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, **1**, 807–820.
- Gonze, D. *et al.* (2005) Spontaneous synchronization of coupled circadian oscillators. *Biophys. J.*, **89**, 120–129.
- Goodwin, B.C. (1965) Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.*, **3**, 425–437.
- Gopalakrishnan, A. *et al.* (2011) Incorporating delayed and infrequent measurements in Extended Kalman Filter based nonlinear state estimation. *J. Process Contr.*, **21**, 119–129.
- Grima, R. (2012) A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.*, **136**, 154105.
- Grima, R. (2015) Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems. *Phys. Rev. E*, **92**, 042124.
- Heron, E.A. *et al.* (2007) Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics*, **23**, 2596–2603.
- Hey, K.L. *et al.* (2015) A stochastic transcriptional switch model for single cell imaging data. *Biostatistics*, **16**, 655–669.
- Jazwinski, A.H. (2007) *Stochastic Processes and Filtering Theory*. Dover Publication Inc., Mineola, NY. [https://books.google.it/books?hl=it&lr=&id=4AqL3vE2J-sC&oi=fnd&pg=PP1&dq=stochastic+processes+and+filtering+theory+&ots=414UF94xaO&sig=p\\_qkoTbC24eA0Cv3g7nF-6tB8rM#v=onepage&q=stochastic%20processes%20and%20filtering%20theory&f=false](https://books.google.it/books?hl=it&lr=&id=4AqL3vE2J-sC&oi=fnd&pg=PP1&dq=stochastic+processes+and+filtering+theory+&ots=414UF94xaO&sig=p_qkoTbC24eA0Cv3g7nF-6tB8rM#v=onepage&q=stochastic%20processes%20and%20filtering%20theory&f=false)
- Jenkins, D.J. *et al.* (2013) A temporal switch model for estimating transcriptional activity in gene expression. *Bioinformatics*, **29**, 1158–1165.
- Julier, S.J. and Uhlmann, J.K. (1997) New extension of the Kalman filter to nonlinear systems. In: *Signal Processing, Sensor Fusion, and Target Recognition VI*. Vol. 3068. International Society for Optics and Photonics, Bellingham, Washington, pp. 182–194.
- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *J. Fluids Eng.*, **82**, 35–45.
- Kim, J.K. and Forger, D.B. (2012) A mechanism for robust circadian timekeeping via stoichiometric balance. *Mol. Syst. Biol.*, **8**, 630.
- Kim, J.K. *et al.* (2014) Molecular mechanisms that regulate the coupled period of the mammalian circadian clock. *Biophys. J.*, **106**, 2071–2081.
- Ko, C.H. *et al.* (2010) Emergence of noise-induced oscillations in the central circadian pacemaker. *PLoS Biol.*, **8**, e1000513.
- Komorowski, M. *et al.* (2009) Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, **10**, 343.
- Korenčić, A. *et al.* (2012) The interplay of cis-regulatory elements rules circadian rhythms in mouse liver. *PLoS One*, **7**, e46835.
- Kulikov, G.Y. and Kulikova, M.V. (2014) Accurate numerical implementation of the continuous-discrete extended Kalman filter. *IEEE Trans. Automat. Contr.*, **59**, 273–279.
- Kurtz, T.G. (1972) The relationship between stochastic and deterministic models for chemical reactions. *J. Chem. Phys.*, **57**, 2976–2978.
- McQuarrie, D.A. (1967) Stochastic approach to chemical kinetics. *J. Appl. Probab.*, **4**, 413–478.
- Monk, N.A. (2003) Oscillatory expression of Hes1, p53, and NF- $\kappa$ B driven by transcriptional time delays. *Curr. Biol.*, **13**, 1409–1413.
- Novák, B. and Tyson, J.J. (2008) Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.*, **9**, 981.
- Relógio, A. *et al.* (2011) Tuning the mammalian circadian clock: robust synergy of two loops. *PLoS Comput. Biol.*, **7**, e1002309.
- Särkkä, S. (2013) *Bayesian Filtering and Smoothing*. Vol. 3. Cambridge University Press, Cambridge.
- Schnoerr, D. *et al.* (2017) Approximation and inference methods for stochastic biochemical kinetics - a tutorial review. *J. Phys A Math. Theor.*, **50**, 093001.
- Singer, H. (2002) Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood versus extended Kalman filter and Itô-Taylor expansion. *J. Comput. Graph. Stat.*, **11**, 972–995.
- Stathopoulos, V. and Girolami, M.A. (2013) Markov Chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Philos. Trans. Royal Stat. Soc. A*, **371**, 20110541.
- Steuer, R. *et al.* (2003) Constructive effects of fluctuations in genetic and biochemical regulatory systems. *Biosystems*, **72**, 241–251.
- Tkačik, G. and Walczak, A.M. (2011) Information transmission in genetic regulatory networks: a review. *J. Phys. Condens. Matter*, **23**, 153102.
- Van Kampen, N.G. (1992) *Stochastic Processes in Physics and Chemistry*. Vol. 1. Elsevier, Amsterdam.
- Wilkinson, D.J. (2012) *Stochastic Modelling for Systems Biology*. Chapman & Hall, London.
- Yamaguchi, S. *et al.* (2003) Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science*, **302**, 1408–1412.
- Zechner, C. *et al.* (2012) Moment-based inference predicts bimodality in transient gene expression. *Proc. Natl. Acad. Sci. USA*, **109**, 8340–8345.
- Zhu, B. *et al.* (2017) A Cell-Autonomous Mammalian 12 hr Clock Coordinates Metabolic and Stress Rhythms. *Cell Metab.*, **25**, 1305–1319.